Validation of the kidney failure risk equation in European CKD patients

Mieke J. Peeters\textsuperscript{1,2}, Arjan D. van Zuilen\textsuperscript{3}, Jan A.J.G. van den Brand\textsuperscript{1}, Michiel L. Bots\textsuperscript{4}, Peter J. Blankestijn\textsuperscript{3} and Jack F.M. Wetzels\textsuperscript{1} for the MASTERPLAN Study Group\textsuperscript{†}

Correspondence and offprint requests to: Mieke J. Peeters; E-mail: M.Peeters@nier.umcn.nl

\textsuperscript{†}See the Acknowledgements section for MASTERPLAN collaborators.

ABSTRACT

Background. Patients with chronic kidney disease (CKD) are at risk for progression to kidney failure. Using data of Canadian CKD patients, Tangri et al. recently developed models to predict the progression of CKD stages 3–5 to kidney failure within 5 years. We validated this kidney failure risk equation (KFRE) in European CKD patients.

Methods. We selected non-transplanted patients with CKD stages 3–5 who participated in the MASTERPLAN study, a randomized controlled trial in patients with CKD. Kidney failure was defined as the initiation of chronic dialysis or kidney transplantation within 5 years. Patients who died before kidney failure were censored. Patients followed for <5 years, who did not develop kidney failure and did not die, were excluded. The 5-year kidney failure risk was predicted using three different models developed by Tangri et al. and compared with the actual kidney failure rate in MASTERPLAN. Model performance was evaluated using the area under the receiver operating characteristic curve (ROC-AUC), the net reclassification index (NRI) and by comparing the observed and predicted rates of kidney failure.

Results. A total of 595 patients were included; 114 developed kidney failure. (Overall observed kidney failure risk in our cohort was 5% lower than in the Canadian validation cohort.) Discrimination of the eight-variable model [including age, sex, estimated glomerular filtration rate (eGFR), albuminuria, calcium, phosphate, bicarbonate, albumin] was similar to that of the four-variable model (including age, sex, eGFR, albuminuria) and the three-variable model (including age, sex, eGFR); ROC-AUCs were 0.89 [95% confidence interval (CI) 0.86–0.92], 0.88 (95% CI 0.85–0.91) and 0.88 (95% CI 0.85–0.92), respectively. Using the NRI, the eight-variable model slightly outperformed the four-variable model (NRI 6.5%) and the three-variable model (NRI 12.4%). The mean differences between the observed and predicted kidney failure risk were −4.0, −7.1 and −7.4% for the eight-, four-, and three-variable model, respectively.

Conclusions. The KFRE accurately predicted the progression to kidney failure in European CKD patients. Discrimination of the three models was similar. Calibration of the eight-variable model was slightly better than that of the simpler models. We question whether this outweighs its added complexity.
INTRODUCTION

Chronic kidney disease (CKD) is a major public health problem worldwide [1]. Patients with CKD are at risk for progression to kidney failure [1, 2], which can be prevented or delayed by early detection and treatment of CKD [3]. However, treatment adds costs and may also have adverse effects. Treatment-related decisions are now often guided by the severity of CKD, based upon estimated glomerular filtration rate (eGFR) and albuminuria. However, these data are considered inadequate to predict the progression to kidney failure [4, 5], neither extrapolation of reciprocal creatinine plots is reliable in predicting the onset of dialysis [6], resulting in treatment delays in those who progress to kidney failure and unnecessary treatment in those who do not progress [5]. Accurate prediction of kidney failure risk could help balance the risks and benefits of treatment, and facilitate decision-making in individual patients. Moreover, accurate risk prediction would support policy makers in dialysis capacity planning.

Until recently, no generalizable prediction model was available. Using data of Canadian CKD patients, Tangri et al. developed models to predict the progression of CKD stages 3–5 to kidney failure within 5 years [5]. The models used clinical parameters and routinely available laboratory data. The most accurate model was an eight-variable model, including age, sex, eGFR, albuminuria, calcium, phosphate, bicarbonate and albumin [5]. The authors reported promising model performance in Canadian CKD patients [5], but the models were not validated in other cohorts. In this study, we validated the kidney failure risk equation (KFRE) in European CKD patients.

MATERIALS AND METHODS

We selected non-transplanted patients with CKD stages 3–5 from the MASTERPLAN (Multifactorial Approach and Superior Treatment Efficacy in Renal Patients with the Aid of Nurse practitioners) study. The MASTERPLAN study is a randomized controlled trial that evaluated the added value of nurse practitioner care in reducing cardiovascular events and attenuating the decline of kidney function in patients with prevalent CKD. Patients were randomized to receive nurse practitioner support added to physician care (intervention group) or physician care alone (control group). Rationale, design and outcome have been published elsewhere [7–9]. Ethics committee approval was obtained as well as written informed consent of all participants. For the analysis of renal endpoints, follow-up was extended until a median follow-up time of 5.7 years.

In the present study, the outcome of interest was kidney failure, defined as the initiation of chronic dialysis or kidney transplantation within 5 years. Patients who died before kidney failure were censored, since the KFRE was developed to predict kidney failure and not mortality [5]. Patients followed for <5 years, who did not develop kidney failure and did not die, were excluded. Patients with CKD stages 3 and 4 were also evaluated separately.

The baseline data of the MASTERPLAN study were used to predict 5-year kidney failure risk by the models that Tangri et al. developed. Next to the most accurate eight-variable model, we also evaluated the four- and three-variable models, to determine whether the benefits of the eight-variable model outweigh its added complexity (Supplementary Figure S1).

Missing data were imputed using an expectation-maximization algorithm [10] (single imputation).

To convert urine albumin-to-creatinine ratio (ACR) to milligram/gram, the milligram/millimole value was multiplied by 8.84. To convert serum calcium to milligram/decilitre, the millimole/litre value was multiplied by 4.01. To convert serum phosphate to milligram/decilitre, the millimole/litre value was multiplied by 3.10. To convert serum albumin to gram/decilitre, the gram/litre value was multiplied by 0.1.

To estimate GFR, the abbreviated MDRD formula [11] was used, as Tangri et al. also used this formula. Calibrated creatinine values were used.

In MASTERPLAN, variable values were not always within the range of allowed values in the KFRE. Therefore, values outside the allowed range were adjusted as dictated by Tangri et al.: if the value was below the minimum value, the minimum value was used. If the value was above the maximum value, the maximum value was used [5].

In our study, we used the KFRE to discriminate CKD patients who would develop kidney failure within 5 years from those who would not. The observed kidney failure rate (on the basis of MASTERPLAN follow-up data) was the reference. Because the outcome of the KFRE can be anywhere between 0 and 100%, a cut-off point, the point on the continuum between ‘normal’ and ‘abnormal’, should be defined. The definition of a cut-off point is arbitrary. At every possible cut-off point, there is a different sensitivity (true-positive rate) and specificity (true-negative rate). In a receiver operating characteristic (ROC) curve, the sensitivity is plotted against 1-specificity (false-positive rate) over a range of cut-off values [12]. In this way, a ROC curve shows the performance of a test, without the definition of one specific cut-off point. ROC curves are particularly valuable ways of comparing alternative tests for the same diagnosis. The overall accuracy of a test can be described as the area under the ROC curve (AUC); the larger the area, the better the test [12]. An area of 0.5 reflects the diagonal (the predictive performance of ‘tossing a coin’) and an area of 1.0 reflects perfect discrimination (sensitivity and specificity both 100%). We compared the ROC-AUCs from the three KFRE models by a method described by Hanley and McNeil [13]. At P-values of <0.05, we considered ROC-AUCs to be different.

Calibration describes the agreement between the predicted risks and observed frequencies of the outcome. We created quintiles of predicted risk and compared the observed and predicted rate of kidney failure. Calibration was tested by the Hosmer–Lemeshow test [14].

Reclassification assesses the movement of patients from one risk category to another when different prediction
models are used. We used the net reclassification improvement (NRI) statistic to quantify reclassification. Risk categories of 0–9.99, 10–19.99 and ≥20% 5-year risk of kidney failure were considered as low, intermediate and high risk, respectively.

Analyses were performed with PASW Statistics 18 (SPSS Inc., Chicago, IL). Comparisons of ROC-AUCs were performed with Stata 11.1 (StataCorp LP, College Station, TX).

RESULTS

Out of the 788 participants in MASTERPLAN, 642 were non-transplanted patients with CKD stage 3–5 at baseline. Of these 642 patients, 114 developed kidney failure within 5 years, 59 patients died within 5 years before kidney failure developed and 422 patients were followed for at least 5 years and did not develop kidney failure. There were 47 patients followed for <5 years, who did not develop kidney failure and did not die. These 47 patients were excluded.

In 37 patients, one or more values from the eight-variable risk prediction model were missing. ACR was imputed in 26 patients, serum albumin in 3 patients, serum phosphate in 2 patients, serum bicarbonate in 10 patients and serum calcium in 2 patients. In 34 patients, one value was imputed, in 2 patients, two values were imputed and in 1 patient, five values were imputed.

Baseline characteristics of the 595 included patients are shown in Table 1. Compared with the Canadian cohorts [5], our patients were younger, more often male, less often diabetic and more often had a history of smoking. Furthermore, our patients had higher diastolic blood pressure, higher serum calcium, lower serum phosphate and higher ACR levels.

Supplementary Table S1 shows the values that were outside the allowed range, and thus adjusted to be used in the KFRE.

The observed versus predicted 5-year kidney failure risk using the three models are shown in Figure 1. All models showed a lower observed than predicted incidence, with only minor differences in calibration between the models. The overall mean differences between the observed and predicted kidney failure risk were: eight-variable model –4.0% (P = 0.03), four-variable model –7.1% (P = 0.05) and three-variable model –7.4% (P = 0.06). Calibration plots are shown in Supplementary Figure S2.

In our study population, ROC-AUCs for the three different models were similar: 0.89 [95% confidence interval (CI) 0.86–0.92], 0.88 (95% CI 0.85–0.91) and 0.88 (95% CI 0.85–0.92) for the eight-, four- and three-variable model, respectively (P = 0.32) (Figure 2). These ROC-AUCs are slightly smaller than the c-statistics reported by Tangri et al.: 0.92, 0.91 and 0.89 for the eight-, four- and three-variable model, respectively [5].

Table 2 shows ROC-AUCs calculated after restricting analysis to patients with CKD stage 3; another separate

---

### Table 1: Baseline characteristics

<table>
<thead>
<tr>
<th>Included patients (n = 595)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60.8 (12.1)</td>
</tr>
<tr>
<td>Male sex</td>
<td>69%</td>
</tr>
<tr>
<td>Caucasian race</td>
<td>92%</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>139 (21)</td>
</tr>
<tr>
<td>Systolic blood pressure &gt;140 (mmHg)</td>
<td>45%</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>80 (11)</td>
</tr>
<tr>
<td>Diastolic blood pressure &gt;90 (mmHg)</td>
<td>15%</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>82.0 (15.1)</td>
</tr>
<tr>
<td>History of diabetes mellitus&lt;sup&gt;a&lt;/sup&gt;</td>
<td>25%</td>
</tr>
<tr>
<td>Prior cardiovascular disease&lt;sup&gt;b&lt;/sup&gt;</td>
<td>30%</td>
</tr>
<tr>
<td>History of current or previous smoking</td>
<td>59%</td>
</tr>
<tr>
<td>Current smoking</td>
<td>20%</td>
</tr>
<tr>
<td>eGFR (mL/min/1.73 m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>33.3 (11.7)</td>
</tr>
<tr>
<td>CKD stage 3 (eGFR 30–59)</td>
<td>57%</td>
</tr>
<tr>
<td>CKD stage 4 (eGFR 15–29)</td>
<td>38%</td>
</tr>
<tr>
<td>CKD stage 5 (eGFR &lt;15)</td>
<td>5%</td>
</tr>
<tr>
<td>Serum calcium (mg/dL)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>9.53 (0.57)</td>
</tr>
<tr>
<td>Serum phosphate (mg/dL)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3.50 (0.79)</td>
</tr>
<tr>
<td>Serum albumin (g/dL)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>4.0 (0.4)</td>
</tr>
<tr>
<td>Serum bicarbonate (mEq/L)</td>
<td>24.8 (3.6)</td>
</tr>
<tr>
<td>Urine ACR (mg/g)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>223 (83–769)</td>
</tr>
<tr>
<td>Urine ACR &lt;30 (mg/g)&lt;sup&gt;g&lt;/sup&gt;</td>
<td>13%</td>
</tr>
<tr>
<td>Urine ACR 30–300 (mg/g)&lt;sup&gt;h&lt;/sup&gt;</td>
<td>46%</td>
</tr>
<tr>
<td>Urine ACR &gt;300 (mg/g)&lt;sup&gt;i&lt;/sup&gt;</td>
<td>41%</td>
</tr>
<tr>
<td>ACEi/ARB use</td>
<td>82%</td>
</tr>
<tr>
<td>Number of antihypertensive drugs</td>
<td>3.0 (1.7)</td>
</tr>
</tbody>
</table>

Values are given as mean (SD), percentage or median (inter-quartile range). eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; ACR, albumin-to-creatinine ratio.

<sup>a</sup>Diabetes mellitus was defined as using blood glucose lowering medication or fasting glucose >7.0 mmol/L.

<sup>b</sup>Cardiovascular disease was defined as myocardial infarction, stroke or vascular intervention.

<sup>c</sup>To convert serum calcium to mmol/L, multiply by 0.25.

<sup>d</sup>To convert serum phosphate to mmol/L, multiply by 0.323.

<sup>e</sup>To convert serum albumin to g/L, multiply by 10.

<sup>f</sup>To convert urine ACR to mg/mmol, multiply by 0.113.
analysis was performed restricted to CKD stage 4. Although the eight-variable model performed better than the four- and three-variable model in CKD stage 3 (P = 0.02), this conclusion is weakened by the low number of kidney failure events in this subgroup.

In Table 3, the summarized results of the reclassification analyses are shown. The complete analyses are shown in Supplementary Table S2. Overall, the NRI of the eight-variable model was 12.4% compared with the three-variable model and 6.5% compared with the four-variable model. Overall, the NRI of the four-variable model was 7.9% compared with the three-variable model.

**DISCUSSION**

In this study, we show that the KFRE accurately predicted the progression to kidney failure in European CKD patients. Calibration of the eight-variable model was slightly better than that of the simpler models. Using the NRI as a measure of reclassification, the eight-variable model was superior. However, the meaning of the NRI is dependent on the clinical relevance of the chosen threshold values [15]. In our NRI calculation, the thresholds were set arbitrarily. Therefore, we are reluctant to build a conclusion merely on the NRI score, and we...
question whether the small advantage of the eight-variable model outweighs its added complexity.

In addition, when discrimination was evaluated using ROC-AUCs (which are particularly useful for investigating the extent to which a biomarker improves the predictive performance of a risk model [15]), performance was similar across all three models.

It is difficult to judge the clinical utility of the predictive models by using ROC-AUCs or NRIs. Therefore, we have tried to provide some more information by calculating test characteristics using a cut-off value of 80% as high-risk predictor and in a separate analysis 20% as low-risk predictor (Supplementary Table S3).

We hypothesized that at very low eGFR, the actual GFR is most important in the prediction of progression of CKD to kidney failure. The additional variables included in the eight-variable model (serum calcium, phosphate, bicarbonate and albumin) may, therefore, play a more important role predicting kidney failure in CKD stage 3 than in the higher CKD stages. This is why we evaluated the KFRE models’ performance separately in CKD stages 3 and 4. (The number of patients with CKD stage 5 was too small to draw a conclusion.)

In CKD stage 3, the eight-variable model indeed seemed slightly more accurate than the four- and three-variable models. However, kidney failure event rate was very low (4%). In CKD stage 4, discrimination was similar for all three models. In line with our hypothesis, other variables appear to have limited prognostic value beyond that of eGFR in patients with CKD stage 4. We therefore suggest that the value of the KFRE is larger in patients with CKD stage 3 compared with CKD stage 4, although a firm conclusion cannot be drawn.

In the Canadian validation cohort, kidney failure rate was higher than in the development cohort [5]. Consequently, the observed kidney failure risk was higher than predicted. In our population, kidney failure rate was also higher than in the development cohort (19 versus 11%). Therefore, also in our population, one would expect the observed kidney failure risk to be higher than predicted; however, the opposite was true (Figure 1, Supplementary Figure S2). This could be due to improved quality of care during the study period [9].

Our population was selected from a trial. This may limit the generalizability of our results. Moreover, the care that our patients received may not be representative for standard practice and may have influenced the outcome of our validation. Another explanation for differences between our results and Tangri’s could be differences in case mix: prevalence of diabetes, nephrological diagnosis and cardiovascular risk. In addition, differences in life style could play a role.

We imputed one or more missing values in 37 patients. This represents clinical practice, because not all patients will have complete data. Furthermore, imputation of missing values is superior to complete case analysis [16]. We excluded patients who were followed for <5 years and did not develop an endpoint. This could have introduced selection bias.

We already described the importance of accurate risk prediction of CKD progression to kidney failure in treatment-related decision-making. Furthermore, it is well known that implementation of guidelines is rather difficult and treatment goals are often not met in renal patients [17–20]. Inclusion of a risk prediction model in an osteoporosis treatment guideline improved guideline adherence [21]. Similarly, the KFRE models could also become important improving guideline adherence in renal patients. Further, the outcome of the prediction model might stimulate patients in their therapeutic compliance.

As Tonelli and Manns [22] already mentioned, another important value of the risk prediction score would be to identify patients who would benefit most from referral. This also includes (early) referral to specialized predialysis or pretransplant care in patients with high risk for kidney failure; alternatively, patients with low risk for kidney failure may be referred back to primary care or general internal medicine care. As in the Canadian cohorts, all patients included in the MASTERPLAN study were already under the care of a nephrologist.

Therefore, the KFRE should also be validated in the general population or a primary care cohort. However, when a secondary care model is validated in a primary care setting, its predictive performance is usually decreased, because distribution of the outcome and predictive factors are different. Updating the model by recalibration, including adjustment of the intercept and relative weights of predictors, should then be considered [23].

In conclusion, the KFRE accurately predicted the progression to kidney failure in European CKD patients. Discrimination of the three models was similar. Calibration of the eight-variable model was slightly better than that of the simpler models. We question whether this outweighs its added complexity.

---

**Table 3: Net reclassification, 5-year kidney failure risk categories 0–9.99% (low risk), 10–19.99% (intermediate risk) and ≥20% (high risk)**

<table>
<thead>
<tr>
<th>Models</th>
<th>NRI kidney failure (%)</th>
<th>NRI non-kidney failure (%)</th>
<th>NRI overall (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-versus three-variable</td>
<td>−4.4</td>
<td>12.3</td>
<td>7.9</td>
<td>0.02</td>
</tr>
<tr>
<td>Eight-versus three-variable</td>
<td>−8.8</td>
<td>21.2</td>
<td>12.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Eight-versus four-variable</td>
<td>−2.6</td>
<td>9.1</td>
<td>6.5</td>
<td>0.03</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY DATA

Supplementary data are available online at http://ndt.oxfordjournals.org.

ACKNOWLEDGEMENTS

MASTERPLAN collaborators: Peter J Blankestijn, Department of Nephrology, University Medical Center Utrecht, Utrecht, The Netherlands; Michel L. Bots, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; Marjolijn van Buren, Department of Internal Medicine, Haga Hospital, The Hague, The Netherlands; Marc AGJ ten Dam, Department of Internal Medicine, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands; Karin AH Kaasjager, Department of Internal Medicine, Rijinstate Hospital, Arnhem, The Netherlands; Gerry Ligtenberg, Dutch Health Care Insurance Board, Diemen, The Netherlands; Yvo W.J. Sijpkens, Department of Nephrology, Leiden University Medical Center, Leiden, The Netherlands; Henk E. Sluiter, Department of Internal Medicine, Deventer Hospital, Deventer, The Netherlands; Peter J.G. van de Ven, Department of Internal Medicine, Maasstad Hospital, Rotterdam, The Netherlands; Gerald Vervoort, Department of Nephrology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; Louis-Jean Vleming, Department of Internal Medicine, Haga Hospital, The Hague, The Netherlands; Jack F.M. Wetzels, Department of Nephrology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; Arjan D. van Zuilen, Department of Nephrology, University Medical Center Utrecht, Utrecht, The Netherlands.

FUNDING

The MASTERPLAN Study was supported by grants from the Dutch Kidney Foundation (Nierstichting Nederland, number PV 01), and the Netherlands Heart Foundation (Nederlandse Hartstichting, number 2003 B261). Unrestricted grants were provided by Amgen, Genzyme, Pfizer and Sanofi-Aventis.

CONFLICT OF INTEREST STATEMENT

None declared.

(See related article by Acedillo et al. The kidney failure risk equation: on the road to being clinically useful? Nephrol Dial Transplant 2013; 28: 1623–1624.)

REFERENCES


8. van Zuilen AD, van der Tweel I, Blankestijn PJ et al. Multifactorial approach and superior treatment efficacy in renal patients with the aid of nurse practitioners. Design of the MASTERPLAN Study [ISRCTN73187232]. Trials 2006; 7: 8


Cholecalciferol in haemodialysis patients: a randomized, double-blind, proof-of-concept and safety study

Pierre Delanaye¹,
Laurent Weekers¹,
Xavier Warling²,
Martial Moonen²,
Nicole Smelten³,
Laurent Médart⁴,
Jean-Marie Krzesinski¹
and Etienne Cavalier⁵

¹Nephrology-Dialysis-Transplantation, University of Liège, CHU Sart Tilman, Liège, Belgium,
²Nephrology-Dialysis, Centre Hospitalier “La Citadelle”, Liège, Belgium,
³Nephrology-Dialysis, Centre Hospitalier “Bois de l’Abbaye”, Seraing, Belgium,
⁴Radiology, Centre Hospitalier “La Citadelle”, Liège, Belgium and
⁵Clinical Chemistry, University of Liège, CHU Sart Tilman, Liège, Belgium

Correspondence and offprint requests to: Pierre Delanaye;
E-mail: pierre_delanaye@yahoo.fr

ABSTRACT

Background. The role of cholecalciferol supplementation in end-stage renal disease (ESRD) patients has been questioned. The objective of this randomized double-blinded study is to assess whether cholecalciferol therapy can increase serum 25-hydroxyvitamin D [25(OH)D] levels in haemodialysed patients and the safety implications of this therapy on certain biological parameters and vascular calcifications score.

Methods. Forty-three haemodialysis patients were randomized to receive placebo or cholecalciferol (25 000 IU) therapy every 2 weeks. The biological parameters, serum calcium, phosphorus, 25(OH)D and parathormone (PTH) levels, were monitored monthly for 12 consecutive months. Vascular calcifications were assessed by lateral X-ray radiography.

Results. At baseline, the mean serum 25(OH)D levels were low and similar in both groups. Thirty patients (16 treated and 14 placebo) completed the study: 11 patients died (5 placebo and 6 treated), 1 patient dropped out and 1 patient was transplanted (both from the placebo group). After 1 year, the percentage of 25(OH)D deficient patients was significantly lower in the treated group. None of the patients developed hypercalcaemia. The PTH levels tended to increase over the study period under placebo and to decrease in the cholecalciferol group. The median changes in PTH levels from baseline to 1 year were statistically different between the two groups [+80

© The Author 2013. Published by Oxford University Press on behalf of ERA-EDTA. All rights reserved.